

**SHORT REPORT FOR SOLUTION**  
**JRS 2012 Data Mining Competition: Topical Classification of Biomedical Research Papers**  
**Kurakin Alexander**

**Registered as:** kurakin

**Team:** Kurakin Alexander, [kuraga333@mail.ru](mailto:kuraga333@mail.ru)

**Members:** Kurakin Alexander, [kuraga333@mail.ru](mailto:kuraga333@mail.ru), CMC faculty (<http://cmc.msu.ru/>, [cmc@cs.msu.su](mailto:cmc@cs.msu.su)) of Moscow State University (<http://msu.ru/>), Moscow, Russian Federation

**Contacts:** e-mail - [kuraga333@mail.ru](mailto:kuraga333@mail.ru)

**Best Result:** 0.515

**Method was used description:**

Linear Classification was used. Shogun Toolbox (<http://www.shogun-toolbox.org/>) and Octave 3.2 (<http://www.octave.org/>) was used. I'll subscribe params of the best solution's algorithm.

I have classified data with Shogun's 'LIBLINEAR\_L2R\_LR' classifier. Normalization constant  $C=0.12$ , class separation threshold  $t=-1.05$ .

Then I filtered topics (with less classifier's result) if number of them was greater then 6.

Then I added topics to answer by the next rule: «Add topic A if topic B was added by classifier and B-topic texts is A-topic not less then 7 times of 10 in train set».

If classifier has refused I set 44-topic. This topic is most popular in train set.